Capacity-Achieving Sequences for the Erasure Channel

Peter Oswald * Amin Shokrollahi †

September 25, 2000

Abstract

This paper starts a systematic study of capacity-achieving sequences of low-density parity-check codes for the erasure channel. We introduce a class $\mathcal{A}$ of analytic functions and develop a procedure to obtain degree distributions for the codes. We show various properties of this class which will help us construct new distributions from old ones. We then study certain types of capacity-achieving sequences and introduce new measures for their optimality. For instance, it turns out that the right-regular sequence is capacity-achieving in a much stronger sense than, e.g., the Tornado sequence. This also explains why numerical optimization techniques tend to favor graphs with only one degree of check nodes. Using our methods, we attack the problem of reducing the fraction of degree 2 variable nodes, which has important practical implications. It turns out that one can produce capacity achieving sequences for which this fraction remains below any constant, albeit at the price of slower convergence to capacity.

*Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA, email: poswald@research.bell-labs.com

†Digital Fountain, Inc, San Francisco, CA 94110, USA, email: amin@digitalfountain.com. Work done while second author was at Bell Laboratories, Lucent Technologies.

1 Introduction

Low-density parity-check codes have attracted a lot of attention lately. Very simple and efficient decoding algorithms and the near capacity performance of the codes with respect to these algorithms have made them one of the most powerful classes of codes known to date. Despite recent advances in the asymptotic analysis of these codes [1, 2, 3], for all nontrivial channels except for the erasure channel it is still unknown whether there exist sequences of these codes that meet the Shannon capacity. The case of the erasure channel is the simplest to analyse, and a thorough understanding of this case seems to be a prerequisite for understanding the more general situation. Moreover, [3] showed that many concepts that were first developed for the erasure channel carry over to the case of other more complicated channels. For this reason, we will start in this paper a systematic study of capacity achieving sequences of low-density codes over the erasure channel.

In [4] the authors introduced a simple algorithm for correcting erasures in a low-density parity-check code. To describe the result, we need some piece of notation. We visualize low-density parity-check codes as bipartite graphs between a set of left nodes called variable nodes and a set of right nodes called check nodes. An edge in this graph is said to have left-degree i if it is connected to a variable node of degree i. Similarly, it has right-degree i if it is connected to a check node of degree i. Let λ_i and ρ_i denote the fraction of edges of left-degree and right-degree i, respectively. A degree distribution for the graph is then the pair (λ, ρ), where $\lambda = \lambda(x) = \sum_i \lambda_i x^{i-1}$ and $\rho = \rho(x) = \sum_i \rho_i x^{i-1}$. The main result of [4] states that their simple recovery algorithm is successful on a random graph with degree distribution (λ, ρ) and initial erasure probability δ if

$$\delta\lambda(1 - \rho(1 - x)) < x \tag{1}$$

on the interval $(0, \delta)$. In this case, we say that the pair (λ, ρ) affords δ. The capacity of the erasure channel with probability δ is given by $1 - \delta$. On the other hand, (λ, ρ) determine the rate of the code as $R = 1 - \int_0^1 \rho(x)\mathrm{d}x / \int_0^1 \lambda(x)\mathrm{d}x$. We will call R the rate of the pair (λ, ρ). The first design goal is thus to produce pairs (λ, ρ) such that the maximum δ that satisfies (1) is very close to $1 - R$. We call a sequence (λ^n, ρ^n) of degree distributions capacity-achieving (c.a. for short) if the maximum δ^n afforded by (λ^n, ρ^n) converges to its uper bound $1 - R$ as n tends to infinity.

How can we produce c.a. distributions (λ^n, ρ^n)? A closer study of two examples of c.a. distributions in the literature is very helpful. For the first sequence, called the *Tornado sequence*, $\lambda^n(x)$ is the initial segment of the series $-\ln(1 - x)$ properly normalized to give $\lambda^n(1) = 1$, and $\rho^n(x)$ is the

initial segment of an exponential $\exp(\mu(1-x))$, where μ is computed from the rate-constraint [4]. For the second sequence, called the *right-regular sequence*, we have $\rho^n(x) = x^n$ and $\lambda^n(x)$ is related to the power series $(1-x)^{1/m}$ for some m [5]. Roughly speaking, in both cases we start with a $f(x)$ represented by a Taylor series with non-negative coefficients on $[0,1]$ and satisfying the normalizations $f(0) = 0$, $f(1) = 1$, for which

$$\mathcal{T}f(x) := 1 - f^{-1}(1-x) \tag{2}$$

has again a converging Taylor series expansion with non-negative coefficients. The existence of the inverse function f^{-1} needed in (2) is automatic from the conditions. We therefore define the set

$$\mathcal{P} := \{f(x) = \sum_1^\infty f_k x^k\ ,\ \ x \in [0,1]\ \mid\ f_k \geq 0, f(1) = 1\}, \tag{3}$$

and the set $\mathcal{A}$ as the maximal subset of $\mathcal{P}$ invariant under the action of $\mathcal{T}$:

$$\mathcal{A} := \{f \in \mathcal{P} \mid \mathcal{T}f \in \mathcal{P}\}. \tag{4}$$

Later in Section 2 we will introduce a general method for obtaining degree distributions from elements of $\mathcal{A}$. The nonlinear operator $\mathcal{T}$ has two interesting properties: first, its square is the identity, and second, it commutes with composition $\circ$ in the sense that $\mathcal{T}(f \circ g) = \mathcal{T}g \circ \mathcal{T}f$. Using these identities, we will be able to construct infinitely many c.a. sequences starting from known ones.

¿From a practical point of view achieving capacity is not sufficient. Suppose that (λ^n, ρ^n) is a sequence of degree distributions of rate R. What we would like to know is how fast, if at all, the maximal $\delta = \delta^n$ afforded by (λ^n, ρ^n) converges to $1-R$ as $n \to \infty$. This problem was studied in [5], where it was shown that if a_r is the average degree of the check nodes ($1/a_r = \int_0^1 \rho(x)\mathrm{d}x$) and $\epsilon := 1 - \delta/(1-R)$, where δ is afforded by (λ, ρ), then $a_r \geq \log(\epsilon)/\log(R)$, or, equivalently, $\epsilon \geq R^{a_r}$. In that paper a sequence of degree distributions is called *asymptotically quasi-optimal* if a_r stays bounded by $\mu \log(1/\epsilon)$ where μ is a constant depending on R. It was shown that the Tornado sequence and the right-regular sequence are both asymptotically quasi-optimal. A much more refined notion of optimality would be to directly compare R^{a_r} and ϵ. We call a sequence of degree distributions *asymptotically optimal* if ϵ/R^{a_r} stays bounded by some constant depending on the rate. Obviously, asymptotical optimality implies asymptotical quasi-optimality. Using this notion, we will show later in Section 3 that the right-regular sequence is asymptotically optimal while the Tornado sequence turns out to be only asymptotically quasi-optimal. This result may explain

why numerical optimzation techniques [7] tend to produce sequences that are close to being right regular. Given that this also holds for other types of channels [3], we conjecture that right regular sequences are fundamentally superior on symmetric channels. Using the composition operation, we will also construct infinitely many sequences of asymptotically quasi-optimal degree distributions. It is proved in [6] that for a c.a. sequence (λ^n, ρ^n) the sequence $\delta^n \lambda^n(1-\rho^n(1-x))-x$ and all its derivatives up to any fixed order k converge uniformly to zero in $[0,1)$ as n goes to infinity. Examining the first derivative, this shows that the product $\delta^n \lambda_2^n (\rho^n)'(1)$ converges to 1 as n goes to infinity. As a result, no c.a. sequence has the property that for all sufficiently large n we have $\lambda_2^n = 0$. On the other hand, from a practical point of view it is advantageous to have as little degree 2 variable nodes as possible, since these are the nodes that get corrected last. In particular, if the fraction of degree 2 nodes is at most $1-R$, then one can construct the graph in such a way that all the redundant symbols fall within the set of degree 2 variable nodes, and one does not need to correct these nodes at all. This greatly accelerates the decoding procedure. Using our techniques, we will construct in Section 4 asymptotically quasi-optimal sequences of degree distributions whose fraction of degree 2 variable nodes is strictly less than $1-R$.

In this extended abstract we have deliberately omitted the proofs, and numerical evidence in support of our theorems. They can be found in the final version of the paper which can be obtained from the authors upon request.

2 The class $\mathcal{A}$

We start with more details on the properties of $\mathcal{A}$ defined by (4). Note that by definition $f \in \mathcal{A}$ implies absolute convergence of the Taylor series of both f and $\mathcal{T}f$. It is straightforward to check that $\mathcal{T}$ is an involution, i.e., $\mathcal{T}^2 f = f$, that if $f \in \mathcal{A}$ then $S_{a,b} f(x) := \frac{f(ax+b)-f(b)}{f(a+b)-f(b)} \in \mathcal{A}$ for any $0 < a \le a+b \le 1$, and, finally, that for any $f \in \mathcal{P}$, we have $\int_0^1 \mathcal{T}f(x)\,dx = \int_0^1 f(x)\,dx$. Further, it is easy to check that each of the families of functions

$$f(x) = x^n, \quad n = 1, 2, \dots, \qquad f(x) = \frac{e^{ax}-1}{e^a - 1}, \quad a > 0, \qquad f(x) = \frac{(1-b)x}{1-bx}, \quad 0 < b < 1, \quad (5)$$

belong to $\mathcal{A}$. The first two families correspond to the right-regular and Tornado sequences, while (5) has the property $\mathcal{T}f = f$; we call that the *self-inverse sequence.* Many more examples can be constructed from these families by using the above properties of $\mathcal{A}$. Below, we will systematically explore sequences of the form $f(x) = \phi(x^n)$, $n = 1, 2, \dots$, generated by composition of some $\phi \in \mathcal{A}$

with the right-regular sequence. More examples can be found in later sections.

What we outline next is how to produce for given $f \in \mathcal{A}$ and $R \in (0,1)$ a suitable pair of left- and right-degree distributions (λ, ρ) of rate R. Roughly speaking, ρ and λ are defined by scaled sections of the Taylor expansion of f and $\mathcal{T}f$, respectively. Throughout the paper, we use the notation $\lfloor t \rfloor$ for the integer part of a real number t, and set $\{t\} = t - \lfloor t \rfloor \in [0,1)$. If $f(x) = \sum_{k \geq 1} f_k x^k \in \mathcal{P}$, then we set

$$T_t f(x) := \sum_{k \leq \lfloor t \rfloor} f_k x^k + \{t\} f_{\lfloor t \rfloor + 1} x^{\lfloor t \rfloor}, \qquad \hat{T}_t f(x) := \frac{T_t f(x)}{T_t f(1)}. \tag{6}$$

for the Taylor polynomial $T_t f$ of degree $\lfloor t \rfloor + 1 > 0$ of $f \in \mathcal{A}$ and its normalization defined by $\hat{T}_t f(1) = 1$, respectively. Our algorithm for computing the pair (λ, ρ) is as follows.

Algorithm 1. *Given $f \in \mathcal{A}$, $R \in (0,1)$, and an integer N, this algorithm computes a pair (λ, ρ) of rate R and a real number δ such that $\rho(x)$ is of degree N and (λ, ρ) affords δ.*

(1) *If $f(x)$ is a polyomial, then set $\rho(x) := f(x)$, otherwise set $\rho(x) := \hat{T}_N f(x)$. Let $a_r := 1/\int_0^1 \rho(x)\mathrm{d}x$.*

(2) *Let $g(x) := \mathcal{T}f(x) = \sum_{k \geq 1} g_k x^k$. Find t such that $\int_0^1 \hat{T}_t g(x)\mathrm{d}x = \frac{1}{(1-R)a_r}$.*

(3) *Set $\delta := T_t g(1)$ and $\lambda(x) := T_t g(x)/\delta$.*

The final version of the paper states clearly what Step (2) means from a computational point of view. It also proves that if $f(x)$ is analytic at 1, and a_r computed in Step (1) of the above algorithm satisfies $(1-R)a_r \geq 2$, then the algorithm correctly computes its output, i.e., the output satisfies the specifications of the algorithm. The running time of the algorithm depends heavily on how long it takes to compute all the coefficients of g_k necessary. This depends ultimately on the function f and the convergence behavior of its Taylor coefficients. For most of our applications, however, we have analytic expressions for the g_k which makes it rather trivial to estimate the right value of t.

3 Convergence speed of c.a. sequences

This section is devoted to estimating how close the performance of the decoding algorithm for the code given by the pair (λ, ρ) is to the capacity of the erasure channel. In other words, we want to study how close the maximal δ afforded by this pair is to $1 - R$ where R is the rate of the pair. We

introduce $\epsilon(\lambda,\rho) := 1 - \frac{\delta(\lambda,\rho)}{1-R}$. As stated in the introduction, it is proved in [5] that $\epsilon(\lambda,\rho) \geq R^{a_r}$ where $1/a_r = \int_0^1 \rho(x)\,dx$. This shows, that in order to have $\epsilon \to 0$ or, equivalently, $1-\delta \to R$, we must have $a_r \to \infty$ for fixed R. Also, since $\epsilon > R^{a_r}$, we may use either

$$\mu(\lambda,\rho) := \frac{a_r \log(R)}{\log(\epsilon(\lambda,\rho))} \qquad \text{or} \qquad \Delta(\lambda,\rho) := \frac{\epsilon(\lambda,\rho)}{R^{a_r}} \tag{7}$$

to quantitatively measure the closeness of capacity and rate of any particular pair (λ,ρ), hence also of c.a. sequences. In refining the definition introduced in [5], we will call a sequence (λ^n,ρ^n) *asymptotically quasi-optimal with constant* $\mu \geq 1$ if $\limsup_{n\to\infty} \mu(\lambda^n,\rho^n) = \mu$. Clearly, the closer μ is to 1 the better it is. In addition, we call a sequence (λ^n,ρ^n) *asymptotically optimal with constant* $\Delta \geq 1$ if $\limsup_{n\to\infty} \Delta(\lambda^n,\rho^n) = \Delta$. Trivially, asymptotically optimal sequences with any Δ are asymptotically quasi-optimal with constant 1. As it turns out these definitions allow to better classify c.a. sequences, and see distinctive differences between, e.g., the Tornado and right-regular sequences.

Let $\phi \in \mathcal{A}$. We start with sequences (λ^n,ρ^n) of fixed rate $R \in (0,1)$ generated from the sequence

$$f^n(x) := \phi(x^n)\,, \qquad n \geq n_0\,, \tag{8}$$

by Algorithm 1. For short, we say that this sequence is generated by ϕ. Since $\phi \in \mathcal{A}$ all f^n satisfy the additional analyticity condition at $x = 1$. If ϕ is a polynomial of degree K then we choose $N = Kn$ and set $\rho^n = f^n$; otherwise $N = N(n)$ is such that $T_N f^n(1) \geq 1/(1 + f^n(R))$.

Theorem 1. *If a sequence (λ^n,ρ^n) of rate R is generated by $\phi \in \mathcal{A}$ as described above, then it is asymptotically quasi-optimal with constant $\mu \leq \mu^\phi := \frac{1}{\int_0^1 \frac{\phi(x)}{x}\mathrm{d}x}$.*

We remark that Theorem 1 provides only an upper bound for the constant μ of quasi-optimality. However, numerical evidence strongly suggests that the bound is rather sharp. The following result addresses the Tornado and the right-regular sequences and gives a more quantitative statement on their asymptotic quasi-optimality previously established in [4, 5].

Theorem 2. (1) *Let (λ^a,ρ^a) be the sequence of degree distributions of rate $R \in (0,1)$ obtained from the family $f^a(x) = (e^{ax}-1)/(e^a-1)$, $a \to \infty$, by Algorithm 1. This sequence is asymptotically quasi-optimal with constant $\mu \leq \mu_{\mathrm{Torn}}(R) := \frac{\ln(R)}{R-1}$.*

(2) *Let (λ^n,ρ^n) be the right-regular sequence of degree distributions of rate $R \in (0,1)$ obtained from the family (5) by Algorithm 1. This sequence is asymptotically optimal with constant $\Delta = e^\gamma = 1.78107241...$, where $\gamma = 0.57721566...$ is the Euler constant.*

In contrast to the bounds in Theorem 1, the bound of Theorem 2(1) is rate-dependent. Note that $\mu_{\mathrm{Torn}}(R) \to \infty$ as $R \to 0$ and $\mu_{\mathrm{Torn}}(R) \to 1$ as $R \to 1$. As a final comment, we mention that the Tornado family does not produce an asymptotically optimal c.a. sequence. Compared with the statements before it and in light of the numerical evidence, Theorem 2(2) exhibits the excellent asymptotic behavior of the right-regular sequence. We do not know of any asymptotically optimal c.a. sequence with a constant $\Delta < e^{\gamma}$.

4 Reduction of degree 2 nodes

In this short section, we concentrate on the asympotic behavior of the fraction of degree 2 variable nodes of the c.a. sequences generated by the family (8). As discussed in the introduction, this fraction should be as small as possible to allow fast decoding. For an arbitrary left-degree distribution λ, it is defined by the coefficient Λ_2 of the Taylor series $\frac{\int_0^t \lambda(x)\,dx}{\int_0^1 \lambda(x)\,dx} = \Lambda_2 t^2 + \Lambda_3 t^3 + \dots$. The stability condition [5, 3] states in the case of the erasure channel that if (λ, ρ) affords δ, then $\delta\lambda_2\rho'(1) \leq 1$, where $\lambda(x) = \lambda_2 x + \mathrm{O}(x^2)$. Hence, we have $\Lambda_2 = \frac{(1-R)a_r\lambda_2}{2} \leq \frac{(1-R)a_r}{2\delta(\lambda,\rho)\rho'(1)}$. If we apply this inequality to a c.a. sequence (λ^n, ρ^n) of rate R, then using $\delta^n = \delta(\lambda^n, \rho^n) \to 1 - R$ we obtain

$$\lim_{n\to\infty} \Lambda_2^n \leq \lim_{n\to\infty} \frac{a_r^n}{2(\rho^n)'(1)}, \qquad a_r^n = \left(\int_0^1 \rho^n(x)\mathrm{d}x\right)^{-1}, \tag{9}$$

if the limits exist (otherwise, the relation holds at least with lim replaced by lim sup or lim inf). Moreover, if the c.a. sequence (λ^n, ρ^n) of rate R is generated by $\phi \in \mathcal{A}$ as explained in Section 3, then we can replace ρ by f and prove the existence of the limits and equality in (9).

Theorem 3. (1) *If the sequence (λ^n, ρ^n) of rate R is generated by $\phi \in \mathcal{A}$, then $\lim_{n\to\infty} \Lambda_2^n = \Lambda^\phi := \frac{\mu^\phi}{2\phi'(1)}$, where $\mu^\phi = (\int_0^1 x^{-1}\phi(x)\,dx)^{-1}$.*

(2) *For any $0 < R < 1$, there exists a c.a. sequence (λ^n, ρ^n) of rate R which is asymptotically quasi-optimal and has a limit $\lim_{n\to\infty} \Lambda_2^n < 1 - R$.*

The proof of part (2) of the above theorem is obtained by studying $\phi(x^n)$ for some $\phi \in \mathcal{A}$. The final version of the paper contains tables that illustrate the quality of the asymptotic estimates in Theorems 1-3.

References

[1] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pp. 249–258, 1998.

[2] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, 2000. To appear.

[3] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, 2000. To appear.

[4] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Practical loss-resilient codes," in *Proceedings of the 29th annual ACM Symposium on Theory of Computing*, pp. 150–159, 1997.

[5] A. Shokrollahi, "New sequences of linear time erasure codes approaching the channel capacity," in *Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes* (M. Fossorier, H. Imai, S. Lin, and A. Poli, eds.), no. 1719 in Lecture Notes in Computer Science, pp. 65–76, 1999.

[6] A. Shokrollahi, *Capacity-achieving sequences*, vol. 123 of *IMA Volumes in Mathematics and its Applications*, pp. 153–166. 2000.

[7] A. Shokrollahi and R. Storn, "Design of efficient erasure codes with differential evolution," in *Proceedings of ISIT'00*, p. 5, 2000.

[8] W. Walter, *Analysis*, vol. 1 und 2. Springer Verlag, Berlin, 1992.